DOCUMENT RESUME

ED 068 495 TM 001 831

AUTHOR Alkin, Marvin C.; Klein, Stephen P.
TITLE Accountability Defined. Evaluating Teachers for Outcome Accountability.
INSTITUTION California Univ., Los Angeles.
PUB DATE May 72
NOTE 11p.
JOURNAL CIT UCLA Evaluation Comment; v3 n3 p1-11 May 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement; Academic Performance; Cost Effectiveness; *Definitions; *Educational Accountability; Education Vouchers; Effective Teaching; Evaluation Criteria; Evaluation Techniques; Institutional Role; National Norms; Performance Contracts; Performance Criteria; Principals; School Responsibility; Standardized Tests; Teacher Evaluation; Teacher Rating; Teaching Skills; Test Reliability; Test Validity

ABSTRACT

In Part I of this report, educational accountability is viewed as being composed of three types. They are: goal, program, and outcome accountability. In addition, three accountability schemes which are considered exemplary of the kinds of proposals presently made and which cover a broad range of accountability types are also discussed. The schemes are: the voucher plan, performance contracting with an external contractor, and performance contracting with a teacher. In Part II, it is suggested that traditional methods of evaluating teachers are not adequate. For example, it is felt that principals' judgments are usually too subjective to provide valid data while standardized tests are often too insensitive to student performance on the specific goals and objectives of a given educational program. It is suggested that student performance on relevant measures be used as the primary basis for a teacher evaluation system. The steps needed for instituting two potentially effective systems--an objectives-based approach to outcome accountability and performance tests--are described along with their advantages and limitations. The difficulties of implementing any kind of an evaluation system are considered as well as how these difficulties might be overcome if the focus of the approach is on the improvement of teacher skills and educational practices.
(Author/JS)

May, 1972 Vol. 3, No. 3

UCLA CSE Evaluation comment

Center for the Study of Evaluation

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 068495

TM 001 831

THE CENTER'S THEORY PROGRAM

The Center for the Study of Evaluation was founded in June, 1966. It is an educational research and development center sponsored by the U.S. Office of Education under the Cooperative Research Act and is the only federally funded Center working exclusively on problems in educational evaluation.

The Center mission – *to improve educational evaluation practice* – is guided by a strategy of concept, tool, and technology development. These activities are conducted simultaneously to maximize development of *new* evaluation technologies and to produce concepts and tools to improve *existing* evaluation systems. With the emphasis on educational decision making, these products are of use in the practical conduct of different kinds of evaluation.

This approach to evaluation is reflected in the Center's organizational structure and the goals of its current activities, which are contained in four programs:

- The Theory Program
- The Program for Research on Objectives-Based Evaluation
- The Higher Education Evaluation Program
- The Evaluation Technologies Program

The Center staff in these programs provides an interdisciplinary team of specialists in education, measurement, sociology, economics, and administration.

This issue stems from the evaluation theory program, which is directed toward the development of more comprehensive models of evaluation. This function is conducted in two related components: general evaluation theory and measurement theory. Staff members are encouraged to address critical evaluation issues through the auspices of the theory program. These "opportunity projects" generally stem from other professional activities of Center staff, and deal with vital evaluation issues. By capitalizing on these opportunities, the Center encourages staff members to expand their ideas for formal presentation.

The first article in this issue was in part stimulated by a request to write a paper for a Phi Delta Kappa symposium on accountability. The second article is adapted from a paper presented to a conference of the California Teachers Association. Given current interest in accountability, it seems reasonable to present these articles to our readers.

The first article offers a theoretical basis for developing systems of accountability. It defines three types of accountability and discusses each. The second article utilizes the theoretical framework of outcome accountability as a starting point for a discussion of practical elements of such a system.

The two articles illustrate the interrelationships between the accountability systems discussed and the Center's conception of evaluation. Moreover, the two articles demonstrate the Center's concern that theory building, in particular evaluation theory building, be constructed primarily as a tool for improving educational practice.

Marvin C. Alkin
Director
Center for the Study of Evaluation

ACCOUNTABILITY DEFINED

Marvin C. Alkin

UNIVERSITY OF CALIFORNIA, LOS ANGELES

It is over simplistic to say that schools are accountable or they are not. Different areas of participation and negotiated responsibility suggest the need to consider different accountability "types." In this article we propose to view accountability as composed of three types: goal accountability, program accountability, and outcome accountability. These derive from an attempt to answer the question, "Who is accountable to whom and for what?"

Introduction

The public has lost faith in educational instituions. Traditional acceptance of educational programs on the basis of their past performance and apparent but unsubstantiated worth is no longer the rule. The public has demanded that schools demonstrate that resources are being utilized "properly." But this has meant far more than mere financial accounting to ensure that funds have not been illegally spent or embezzled. What is demanded instead is that schools demonstrate that the outcomes they are producing are worth the dollar investment provided by communities. In short, what has been called for is a system of "educational accountability."

But educational accountability is very much like other abstract virtues such as patriotism and truthfulness which are universally acknowledged but not amenable to facile description. Lack of adequate description has been one of the major shortcomings of accountability. The reader investigating the subject for the first time becomes immediately innundated with a plethora of views, schemes, mechanisms and, for that matter, a multitude of definitions.

To say that discussion of accountability has been confusing and that definitions of accountability have been amorphous and imprecise is to understate the problem. Barro (1970) says that the basic premise of accountability is that "profesional education should be held responsible for educational outcomes—for what children learn." Many teachers and teacher organizations have a negative connotation such as "it is for punishment." Some school administrators feel accountability can be used to eliminate some of the "deadwood" in teaching. Boards of trustees frequently feel the same way about eliminating the "deadwood" and "overstaffing" in administration. Some economists view accountability as a panacean information system which will cure educational ills by ensuring the wisest allocation of scarce resources. To many people, then,

Marvin C. Alkin, Director of CSE and the Theory Program

accountabilty is the answer. It is "in," however, in a variety of ways for different kinds of proponents.

Popham (1970) asserts that "educational accountability means that the instructional system designer takes responsibility for achieving the kinds of instructional objectives which are previously explicated." Lopez (1970) casts the definition in a social context: "Accountabiilty refers to the process of expecting each member of an organization to answer to someone for doing specific things according to specific plans and against certain timetables to accomplish tangible performance results." Lieberman (1970) asserts that the objective of accountability is to relate results to resources and efforts in ways that are useful for policy making, resource allocation, or compensation.

Smith (1971) suggests three kinds of accountability; program accountability, process accountability, and fiscal accountability. *Program* accountability is concerned with the quality of the work carried on and whether or not it met the goals set for it. *Process* accountability asks whether the procedures used to perform the research (teaching) were adequate in terms of the time and effort spent on the work, and whether the experiments (lessons) were carried out as promised. *Fiscal* accountability has to do with whether items purchased were used for the project, program, etc.

Lessinger (1970) has said, "accountability is the product of a process; at its most basic level, it means that an agent, public or private, entering into a contractual agreement to perform a service will be answerable for performing according to agreed-upon terms, within an established time period, and with a stipulated use of resources and performance standards."

Definition

In this paper we will tentatively settle on a definition of accountability as:

> *Accountability is a negotiated relationship in which the participants agree in advance to accept specified rewards and costs on the basis of evaluation findings as to the attainment of specified ends.*

The essence of this definition is that a negotiated relationship exists in which each of the participants agree in advance as to the criteria (evaluation findings) that will be used to determine acceptability. Furthermore, the level of attainment on these criteria in order to achieve acceptability is pre-specified. Finally, the negotiants stipulate a set of rewards and penalties that will attach to compliance/non-compliance.

At the heart of all of the above elements is the concept of "negotiation." Negotiation, for example, is suggested in the kind of dialogue which leads to mutual acceptance of a position, or in the acceptance of a negotiated, specified end. Negotiation frequently involves the allowable constraints, such as the students to be worked with and the instructional materials to be utilized. One major form of negotiated relationship, although not the only one, is the written contract. A contractual agreement will specify the locus of problem solving and areas of responsibility between the negotiants. To establish these relationships, a contract will provide with utmost explicitness and clarity, the following:

- A set of stated constraints
- The negotiated ends in light of the constraints
- Designation of responsibility in terms of who is responsible for what, to whom, and when
- Criteria for judging attainment of ends
- Specification of the rewards and costs to include payment and penalty schedules

Before such contractual explicitness can be achieved in terms of relationships betwen negotiants in a system of accountability, we must first address some contextual considerations and discuss the major segments within that context. Without such specification it is virtually impossible to adequately address the locus of problem solving and areas of responsibility in any manageable form.

We view the three major segments of the accountability context as: (1) goals and objectives, (2) programs, (3) program outcomes. A system of accountability can be functional only in those educational institutions which have clearly defined goals and objectives. These goals and objectives derive from interactions with various constituencies whose views are thought to be relevant and whose priorities are reflected in the specified outcomes. For these objectives, which in turn are related to the broader goals, there are specific, clearly defined, and validated instructional programs or strategies. The instructional programs or strategies have been validated to the extent that there are specific product specifications demonstrating the success of the programs relative to the stated objectives of the program for various kinds of population groups, one

of whom is the group for which it will be employed. A further element of this context is a specific procedure for measuring the program's outcome in terms of the stipulated objectives. To the extent that the school context approaches such a rational effort, it is possible to have an accountability system.

Accountability Types

Part of the differing conceptions of accountability undoubtedly stem from our insistence that accountability is unidimensional. It is over simplistic to say that schools are accountable or they are not. For each area of the context there can be different role participants. Different areas of participation and responsibility suggest the need to consider different accountability "types;" the three components outlined above suggest that there are perhaps three types of accountability.

We propose to designate the three types of accountability as *goal accountability*, *program accountability*, and *outcome accountability*. These three accountability types derive from an attempt to answer the question, "Who is accountable to whom and for what?" When this question is considered with respect to the context areas listed above, we note that different participants are involved on various occasions.

The first area to be considered is *goal accountability*. School boards are accountable (or should be) to the public for everything that they do. But the foundation of this accountability relationship is in educational goals. School boards are accountable to the public for the proper selection of goals. After all, school boards are legally supposed to function as the lay group expressing the desires and wishes of a broader constituency as to what should be the goals and objectives of the educational program. This determination is clearly within the domain of the public's review responsibility. In goal accountability, school boards are accountable to the public for ensuring that the proper goals and objectives are being pursued in the school program.

After goals and objectives are selected, responsibility rests somewhere for the selection of instructional strategies deemed most effective for achieving the stipulated goals and objectives. This responsibility for *program accountability* rests generally with the school administration and other school personnel designated by administration. If we conceive of the teachers as being program operators and intend to hold them accountable for the outcomes of their activities, then clearly they may only be held accountable within the constraints of the programs with which they have been provided. The responsibility for program accountability rests with administrators and other members of the professional staff engaged in the process of program selection, modification, and adoption.

In program accountability, these administrators and other district personnel, though again ultimately responsible to the public, are specifically accountable to the school board for maintaining a program which is appropriate for meeting a set of stipulated objectives. We cannot hold a machine operator responsible for his products until we have demonstrated that the machine he has been provided with has the capability for producing that outcome. We cannot expect a printing press operator to produce 100 copies a minute on a machine whose maximum output is 50 copies per minute. We cannot expect a racetrack driver to push 300 miles an hour out of an automobile whose limit is far below that standard.

If we are to follow this line of argument to its logical conclusion, then clearly, producers of program components (let us refer to these as instructional products) must be held accountable for the products they produce. This is an area of accountability about which we have heard very little. While there is considerable demand that the classroom teacher be accountable, where is the outcry for accountability on the part of textbook producers? Who demands that producers of film strips, films, and supplemental materials present the specifications of their products in terms of outcomes that may be anticipated?

As part of the standards implied in program accountability, a demand should be placed on those to be held accountable for instructional programs that the producible program outcomes be stipulated in terms of the various sets of constraints and the varying inputs that might be encountered. That is, one cannot merely stipulate, without a considerable loss of accuracy in description, that a given product will produce objectives A, B, and C at a given level of achievement. It is also necessary to indicate what the expectations would be for different characteristics of student inputs (for different student groups). This is similar to the example previously discussed in which a printing press operator might be expected to produce 100 copies a minute on a given machine. It is important in that example to consider such things as the quality and weight of the paper to be used, color of the ink, type of master plate, etc. In the race car example it is necessary to be aware of the performance standards for different kinds of roads and weather conditions. Similarly, in educational accountability it is important to have an indication of the performance standards for each program in terms of a variety of input constraints.

With respect to program accountability a difficult and confused area is the role of teachers in, and as a part of, instructional programs. The confusion is amply demonstrated by the diverse views as to what is meant by "teacher accountability." For example, there are those who maintain that teacher accountability is determined on the basis of input standards for teachers. That is, a teacher is accountable if he demonstrates that he is an able teacher in terms of his ability to teach* and by satisfactory application of his skills in terms of the amount of effort put forth on his job. This view of the teacher's role basically considers the teacher as a program component, a part of the instructional program. Under such a definition of teacher accountability one merely looks at teachers as a potential input or program component. Here teacher ac-

*See the discussion on performance testing in the article which follows.

countability is judged in the same way that a textbook, film, or a film strip is considered; the accountability task under such a viewpoint is to ensure the quality of the teacher input. Thus, we may use teacher performance tests as a basis for determining whether teachers participating in the program meet a standard of accountability in terms of their ability to teach.

Within this same definition of the role of teachers, but beyond certification of teacher-input quality, there is a further consideration of the accountability task. This area of accountability responsibility relates to the proper utilization of teacher input. That is, accountability requirements demand that there be an assurance that the inputs (teachers) are working an appropriate number of hours using those skills considered to be appropriate. The notion of teachers *as part of* instructional programs requires accountability examination in terms of input and process evaluation.

A second view of teacher accountability is one in which the teacher is urged to be responsible for the quality of student outputs. In this view the teacher is considered as an instructional manager utilizing a program whose capabilities have already been determined. Here the teacher is held responsible for the outcomes of his management of that program. This type of accountability we will refer to as *outcome accountability*. In this framework we do not question the teacher on process characteristics such as score on a teacher performance test, or the amount of time spent in the classroom, or the processes used. Instead, what is said is "Here is a program whose capabilities have been demonstrated. Show that you are able to produce student outcomes of the desired type and standards using that program."*

In *outcome accountability*, an instructional leader (usually a teacher) is accountable to administration for specified pupil outcomes thought to be a function of teacher management of the instructional program. That is, a teacher manages an instructional program which has certain product capabilities; the job is to determine whether the teacher has managed the program in such a way as to achieve standards or criteria that might be expected from the program.

We have previously said, however, that teachers may only be held accountable within the constraints of the program with which they have been provided. There are those who would maintain, however, that the accountability concern should not focus upon these constraints since the teacher, to a great extent, *is* the program. In this light, in terms of financial outlay for program operation, those costs incurred directly by the teacher amount to the major portion of the available budget. Further, there is sufficient evidence that program constraints have minimal impact upon student outcomes. One would not deny that the teacher incurs the greatest amount of cost in program operation or that program constraints have only a small effect. Yet teachers do work with constraints, such as type of students, kind of text, size of classroom, etc. Though the effects of these constraints may be small, they do, to varying degrees, affect the management of the program and to that extent must be considered in outcome accountability.

*In the article which follows we discuss a comparative procedure for setting standards in outcome accountability using programs for which no standards exist. The necessity for discussing that procedure bears ample evidence to the sorry state of currently validated programs.

Accountability Types: Summary

We have already discussed three major accountability types (goal, program, and outcome) and have indicated a response for each relative to the question, "Who is accountable to whom and for what?" A summary description of each of these types, along with three sets of factors, is presented in Chart 1: (1) Who is accountable – the specific individual or group bearing the responsibility, (2) To whom – the individual or group demanding accountability, (3) For what – specific tasks required.

CHART 1:
Accountabiilty Types

	Who is Accountable	To Whom (Primary Responsibility)	For What
Goal Accountability	School Board	Public	Goal & Objective Selection
Program Accountability	School District Management	School Board	Development and/or Selection of Instructional Programs Appropriate for Stated Objectives
Outcome Accountability	Instructional Manager (i.e., Teacher)	School District Management	Producing Program Outcomes Consistent with Pre-Selected Objectives at a Performance Standard Appropriate for the Instructional Program

Implications of Various Accountability Schemes

A number of schemes have been noted in the literature for achieving greater accountability in schools. Many of these, such as the voucher plan or performance contracting, have been thought of as almost synonymous with accountability. It is important to recognize, however, that these accountability schemes cannot be understood properly without considering to which accountability types they are addressed (e.g., goal accountability, program accountability, outcome accountability) and how they fit within the accountability context previously described.

We will consider three accountability schemes that are fairly exemplary of the kinds of proposals presently made and which cover a broad range of accountability types. These three schemes are the voucher plan, performance contracting with an external contractor, and performance contracting with a teacher.

Under the voucher plan the school passes on the responsibility for all three kinds of accountability. By giving fund grants directly to parents for their expenditure on a program of their own choosing, the school is in essence relieving itself of the full accountability responsibility. No longer must schools be accountable for goals, because parents with funds in hand will choose educational institutions or programs having goals compatible with their

preferences. By such a choice parents and not public schools will be holding their own contractor responsible for both program and outcome accountability. Thus the voucher plan represents a complete irresponsibility on the part of public schools in terms of accountability.

Under performance contracting with an external contractor, while the school retains the responsibility for goal accountability, the contractor becomes responsible for program and outcome accountability. In essence, appropriate goals have been decided upon for a program; the school has consulted with various constituencies about the relevance of various goal areas and has selected a goal or set of goals most worthy of consideration. The external performance contractor is held responsible for the creation of a program to meet these goals as well as for the implementation and management of that program. That is, the external contractor must show both program and outcome accountability. If the community complains about the program and feels that the schools have not achieved the desired outcomes, it is the responsibility of the external contractor; he has obviously failed to do his job. The only way the school can be held accountable is if there is criticism that the goals being pursued are incorrect or inappropriate.

In a system of performance contracting in which the teacher rather than the external contractor is the instructional manager, the school delegates the responsibility only of outcome accountability. That is, the goals have been determined within the school; the program has been determined within the school, including a specification of its capabilities, and the teacher as an instructional manager is to be held accountable for program outcomes. If the teacher is unable to attain educational outcomes equal to a prespecified standard, and that standard is considered appropriate for the given program and students, then it is the teacher who is held accountable. On the other hand, if there is a question about the adequacy of the program itself for achieving the specified goals and objectives, then the school itself (or the school administration) is found short on the accountability criteria.

What we have demonstrated is that there are three types of accountability and there are various schemes that have been presented whereby different agencies or individuals take the responsibility for various types of accountability. In developing a total accountability program, apparently the first decision to be made is the locus of the responsibility for each of the three accountability types.

References

Barro, S. M. An approach to developing accountability measures for the public schools. *Phi Delta Kappan*, 1970, 52, 196-205.

Lessinger, L. Engineering accountability for results in public education. *Phi Delta Kappan*, 1970, 52, 217-225.

Lieberman, M. An overview of accountability. *Phi Delta Kappan*, 1970, 52, 194-195.

Lopez, F. M. Accountability in education. *Phi Delta Kappan*, 1970, 52, 231-235.

Popham, W. J. Instructional objectives exchange, 1960-1970. CSE Reprint No. 19, 1970. Center for the Study of Evaluation, University of California, Los Angeles.

Smith, B. L. R. Accountability and independence in the contract state. In Smith, B. L. R. (Ed.) *The dilemma of accountability in modern government*. New York: 1971, P. 29.

EVALUATING TEACHERS FOR OUTCOME ACCOUNTABILITY*

Stephen P. Klein **Marvin C. Alkin**

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Traditional methods of evaluating teachers are not adequate. Principals' judgments are usually too subjective to provide valid data while standardized tests are often too insensitive to student performance on the specific goals and objectives of a given educational program. This paper suggests that student performance on relevant measures be used as the primary basis for a teacher evaluation system. The steps needed for instituting two potentially effective systems are described along with the advantages and limitations of both systems. The procedures that could be employed by a school district for analyzing and reporting the results relative to student input characteristics are also discussed. Finally, the difficulties of implementing any kind of an evaluation system are considered as well as how these difficulties might be overcome if the focus of the approach is on the improvement of teacher skills and educational practices.

The current emphasis on evaluation and accountability in education has resulted in a number of states passing laws to make them mandatory in one form or another. In California, for example, an accountability law was recently passed requiring the evaluation of teachers in terms of their students' performance. Although we support the rationale underlying such mandates, we often find it discouraging to see how they are worded or implemented. Frequently federal or state governments mandate laws prematurely and with insufficient lead time. Such action puts a severe burden on school personnel who may not be familiar with the issues and methods associated with developing effective evaluation systems. This, in turn, has led to professional evaluators being besieged with requests for

*Based upon a speech presented at the California Teachers Association 21st Annual Good Teaching Conference, January 28, 1972, Los Angeles, California.

Stephen P. Klein, Co-Director of the Evaluation Technologies Program

advice on how to develop such systems so that they are professionally satisfactory and conform to both the letter and the spirit of the law. These requests usually take the form of questions such as "We have a Title III grant to improve student reading and attitudes; how should we evaluate this project?" or "We want to have a teacher-improvement and evaluation system; how should we set it up?"

This paper will focus on the kinds of general advice we would give to answer one facet of the latter question: How should a school set up a system to hold teachers accountable for student outcomes?

By selecting this topic for discussion, we are not addressing the question of whether or not such systems should be developed or, if they are, whether it is also imperative to develop principal, superintendent, and school board accountability systems along with the system for teachers. Those who wish to debate these issues may arm themselves with the preceding paper by Alkin. However, since teacher-evaluation systems are a reality, it is better to have good ones than poor ones. Furthermore, if a school uses a good teacher-accountability system, the quality of education being offered is likely to improve. The rationale to support this contention will be presented later, but first we will consider what a good system should look like.

Requirements Of A Good Teacher-Evaluation System

One way of describing what a good system should look like is to consider what it *should not* look like. First, it *should not* require *subjective* judgments by principals or panels on whether a teacher is performing competently. A good evaluation system should emphasize *objective* assessments of teacher performance. Thus, the common approach of having principals observe and rate teacher performance is not acceptable since it is too open to individual biases. Further, what one principal believes will constitute an effective teacher may not be too highly related to what another principal thinks nor is either of these two subjective judgments necessarily correlated with actual *student* performance. Because of this potential lack of a strong relationship between subjective assessments of teacher quality and demonstrated pupil performance, subjective judgments are likely to be a very poor basis for a good accountability system.

One of the first important features of a good teacher evaluation system, then, is that it be *objective*. Some school districts and state departments of education have sought to achieve such objectivity by relying on nationally-normed standardized tests of student ability and knowledge. The logic behind this approach is that if a teacher does his or her job well, then that teacher's students should learn more than the students of a teacher who is not effective. This seems reasonable, especially if one controls for important factors out of the teacher's control but which still might influence pupil scores. For example, it would be appropriate to compare teachers on the basis of their students' performance if one adjusted the measure of that performance for such factors as the students' previous skills and knowledge. Thus, with the proper controls on certain factors, evaluating teachers on the basis of their students' performance seems like a fair and objective approach.

Unfortunately, the practice of using nationally-normed standardized tests often violates the spirit of this logic. There are several reasons for this, but perhaps the most important is that such measures may be insensitive to the kinds of skills, knowledge, and attitudes that teachers are trying to transmit to their pupils. Nationally-normed tests provide only a single, global score on very general objectives that may have been combined in some very strange ways. These measures may also fail to assess certain objectives considered to be especially important in a given school and these objectives may be among those on which a teacher is devoting most of his class time (Klein, 1970; 1971). Therefore, the use of most nationally-normed standardized tests to assess a given teacher's performance would be analogous to using a bathroom scale to determine how many stamps to put on a letter. A teacher could be very effective and make an important impact on his or her students' performance, but that influence would not register on the measuring scale of nationally-normed tests because such instruments are simply not sensitive enough for the job.

So far we have disqualified one common base for a teacher-evaluation system – ratings from personal contact and observations – and have discussed the possible shortcomings of a second method – nationally-normed standardized measures. In discussing these two kinds of criteria we have mentioned some characteristics that should be considered for a good system. For example, the system should be objective and fair to all the teachers who are going to be evaluated by it. There must, therefore, be some means of adjusting for factors that may influence student performance but over which the teacher has no control. These factors range from prescribed instructional materials (and whether or not they arrive on time) to controlling for

students with different kinds of ability, socio-economic backgrounds, and cultures. Alkin has elaborated such constraints on the teacher in the discussion of program accountability in the opening article. Secondly, the basis for this system should be sensitive to the educational goals and objectives that the school is trying to achieve. It is senseless to say that one teacher is competent and another is not when the basis for this evaluation is how well each of them can teach students to do something which is irrelevant to the school's goals.

Objectives-based Approach To Outcome Accountability

One method of evaluating teachers for outcome accountability that meets the foregoing criteria involves the use of a set of tests or other devices to assess pupil performance on the *particular objectives* with which the school is most concerned. This approach is called "objectives-based evaluation." It usually takes the form of selecting a set of important objectives, constructing short tests to measure each of these objectives, and then administering the tests to all the pupils for whom the objectives are intended. The performance of teachers who are operating under the same conditions can then be compared. One never knows what a legislator is thinking when he drafts a bill, but it was probably the intent of the California legislators to use an objectives-based evaluation system when, in passing their teacher-evaluation law, they said:

> It is the intent of the Legislature to establish a uniform system of evaluation and assessment of the performance of certificated personnel within each school district of the state. The system shall involve the development and adoption by each school district of objective evaluation and assessment guidelines.*

This sounds good, but can it be implemented? First, one must determine what objectives are considered to be most important. To our knowledge, procedures for effectively and economically determining the most important objectives within each district have been developed and implemented statewide in at least one state (Klein, 1972). Thus, it is reasonable to assume that this might eventually be done in districts in other states which are adopting accountability procedures.

The second step in implementing an objectives-based accountability system involves selecting and/or constructing measures to assess student performance on the important objectives. Selecting tests is, of course, a lot easier than constructing them; and books such as the CSE Pre-School/Kindergarten Test Evaluations (1971) and Elementary School Test Evaluations (1970) can be used to facilitate this process *if* there are existing published measures that overlap well with the district's objectives. The construction of measures to assess student performance, on the other hand, especially on objectives involving student attitudes, is a very costly undertaking and not likely to be supportable by each individual school district. It is also rather inefficient since many districts will have essentially the same objectives and, thus, there would be an unnecessary duplication of effort spent on test construction. A state department of education could, therefore, make an important contribution to setting up an accountability system by coordinating and/or supporting the development of the necessary objectives-based measures.

The third step in this process, the administration, scoring, and analysis of the data, could also be done much more efficiently if it were supervised by one central agency. To help ensure unbiased and confidential reports of results this agency might even be a private firm. Such an agency might also handle some of the inherent problems associated with objectives-based systems. One problem, for example, is the sheer number of objectives on which pupils might be assessed if a district wanted to evaluate every teacher's performance on all the objectives that were judged to be important for each teacher's pupils. This might require so much testing time that little would be left for instruction. Alternatively, to say in advance that only a certain group of important objectives will be assessed might encourage some teachers to ignore the other important objectives and thereby penalize those teachers who are conscientious about their profession and who treat all important objectives. In order to alleviate these problems, it has been suggested that when an objectives-based system is employed, it should also involve systematic sampling of students and objectives. This, in turn, will minimize testing time and costs.

Performance Tests

Another approach which has been suggested for establishing a fair and objective basis for a teacher-evaluation system is called "performance tests" (Popham, 1971 a, b). This approach, analogous to the idea of a job sample, is designed to be more efficient than a total objectives-based system and involves selecting a few relevant objectives and constructing tests to measure student achievement of them. The objectives chosen for this purpose should deal with a relatively small but important unit of the curriculum in which the students have had *no previous instruction.* The next step is to assign students to teachers randomly or by means of fair matching techniques so that student characteristics and other factors beyond the teacher's control are counterbalanced among the teachers who are to be evaluated. The teachers are then given a fixed amount of time to teach these objectives and, at the end of that period, student performance is assessed. One assumption underlying this approach is that "teaching ability" is a general characteristic and not limited to just certain kinds of objectives. *Thus, how well a teacher's students do on a series of performance tests is presumed to correlate fairly well with how that teacher's students do on tests to measure end-of-year kinds of objectives.*

The use of performance tests in teacher accountability systems is quite new. There is not yet sufficient data to determine whether these job samples will really reflect teacher proficiency on more than just simple short-term ob-

*Article 5.5, Section 13485. Evaluation and Assessment of Performance of Certificated Employees. California Legislature, 1972.

jectives, but hopes are high that they will. One problem to be faced in the use of teacher proficiency tests is whether a test of teaching ability is a fair criterion or whether the more relevant dimension is teacher achievement. That is, one must view a teacher performance test as a kind of aptitude test rather than achievement test. This has led to the suggestion that teacher performance tests might be used in conjunction with objectives-based evaluation systems to obtain a less costly technique that is relatively easy to use. The procedure would require performing periodic statistical analyses demonstrating the relationships between scores on teacher performance tests and larger batteries of objectives-based measures. If the results were satisfactory, then teacher performance tests could be used as a reasonable proxy for end-of-year outcome measures. At this time, however, using teacher performance tests in this way is an unproven technique and caution is advised.

Setting Standards

No matter what method is chosen, if we are concerned about *judging* teacher performance, then standards must be set. This setting of standards illustrates how the term "evaluation" differs from "assessment" or "measurement" and it is important at this point to specify the nature of this difference. The term "assessment" is used to describe the collection and tabulation of such data as student scores on a test. The word "evaluation" includes assessment but goes beyond that to include a *judgment* of the quality of the obtained measurement. Thus, one could assess a teacher's performance in terms of his or her students' test scores; but to evaluate whether or not that performance is satisfactory one must also have a set of standards against which to judge the quality of that performance. One must ask the question, therefore, for an individual student, whether 75% is acceptable or is 99% needed? Obviously, a host of other kinds of standards or frames of reference might be employed. If one wishes to use the measurement of student performance as a means for judging the quality of teacher effectiveness, then one must set *some* standard against which to evaluate whether or not an individual teacher's performance is acceptable.

There are, of course, many different kinds of standards one might wish to employ. For example, one might set an arbitrary score for the class average. A different kind of standard would involve a comparison of a teacher's effectiveness in improving student performance relative to some norm group, such as students of other teachers. Another approach assumes that students should perform better if they are taught by a professional and qualified teacher rather than by someone who is not a credentialled teacher. Thus, a teacher's effectiveness might be judged in terms of whether his or her students' performance was more like the performance of students taught by a person with or without a credential. It should be noted, however, that Popham (1971b) investigated the utility of this approach and found the results somewhat disconcerting. The reason for his consternation was that he could find *no difference* in the performance of students who were taught by credentialled teachers versus those taught by people off the street. The students in both groups improved equally.[1] It appears, therefore, that if comparisons are to be made to some norm group rather than an absolute standard of performance, then this norm should probably be the performance of pupils of other teachers.

How one should make such comparisons fairly is also an important issue for an accountability system. For example, if one simply looks at the class average, then one ignores the possibility that a high or low average might have been due to just a few extreme cases. Lindman (1968) has suggested, therefore, a technique to see whether a teacher's class improves in performance uniformly or whether the observed end-of-year average score was a function of something happening (or not happening) to certain subgroups within the class (such as those with high, low, or medium ability).

A second problem in the use of a norm group against which to evaluate teacher performance is that pupils are not comparable across teachers. A teacher with bright students should obtain a higher level of performance from these students than that same teacher would with students who were less able. One way around this problem might be to construct different norms for different kinds of students such as those falling at different performance levels on a statewide or district-wide test and for groups using different sets of instructional materials. If other input variables, such as the students' socioeconomic status, were also to be considered, then one would need a very large number of categories and/or advanced statistical grouping techniques such as discriminant function analysis. In any event, the number of teachers in a given district with a sufficient number of pupils in even one category for a given grade (or age) level would probably be so small as to preclude any worthwhile analysis within that district. In short, the norm against which comparisons were to be made would be non-existent. This situation has led a number of researchers, such as Barro (1970), to suggest the use of a technique called regression analysis. The essential features of this approach as it might be applied to an accountability system for a single district are as follows:

1. Administer a pretest to all the students within a given grade or age level. This test should assess each student's performance on all or a good sample of the relevant objectives for students at that level although it is not necessary to have separate scores for each objective. Thus, one might either use a nationally-normed standardized test if it

[1] Similar results have been reported in connection with the Office of Economic Opportunity's study of performance contracting (OEO, 1972). In this experiment, the performance of pupils in regular classrooms with credentialled teachers was compared to that of comparable students receiving special instruction under diverse kinds of conditions. The instruction given to the experimental group ranged from the use of aides with only a few days training to master teachers employing incentives and the most advanced educational technology. OEO reported that there were no significant differences among these approaches! On the other hand, McNeil (1972) has found that students taught by more experienced teachers tend to do better than those taught by teacher trainees.

matches the district's objectives, or construct a measure specifically for the objectives in question. Such a test should not take more than one or, at the most, two hours of testing time.

2. At the end of the year, administer a posttest covering the same objectives that were assessed with the pretest. It would probably be a good idea to use a different set of items for the two tests, however, so as to minimize potential biases.

3. Plot the two sets of scores (pretest and posttest) for each pupil within a given grade (or age) level. The pupil's teacher should also be identified in this process. The scores of five pupils in each of three classes have been plotted in figure 1 to illustrate this procedure.

4. Fit a line among these points on the plot that would represent the average or typical relationship between pretest and posttest scores. The statistical procedure called regression analysis can be used for this purpose.

5. Inspect the results in terms of whether a teacher's class tends to fall above or below the line of expected performance as well as whether the average class performance tends to be above or below this line. Table 1 is an example of how these results might be summarized. An examination of this table reveals that although Teacher A's class had relatively poor pretest scores, they gained more in relation to their starting position than did the students in either Teacher B's class or Teacher C's class. Five students per teacher is, of course, an insufficient number on which to base sound comparative judgments using this procedure and this number was used only for the purposes of illustration. One would need at least 20 or more students in order to get a stable estimate of how well a teacher's students did relative to the typical performance of students at the same grade (or age) level.

TABLE 1

	Teacher A	Teacher B	Teacher C
Average Pretest Score	35	50	65
Average Posttest Score	57	65	73
Expected Posttest Score	50	65	80
Difference between expected & actual posttest scores	7	0	-7
Percent of pupils who are:			
Above Expectancy	80%	40%	20%
At Expectancy	0	40%	0
Below Expectancy	20%	40%	80%

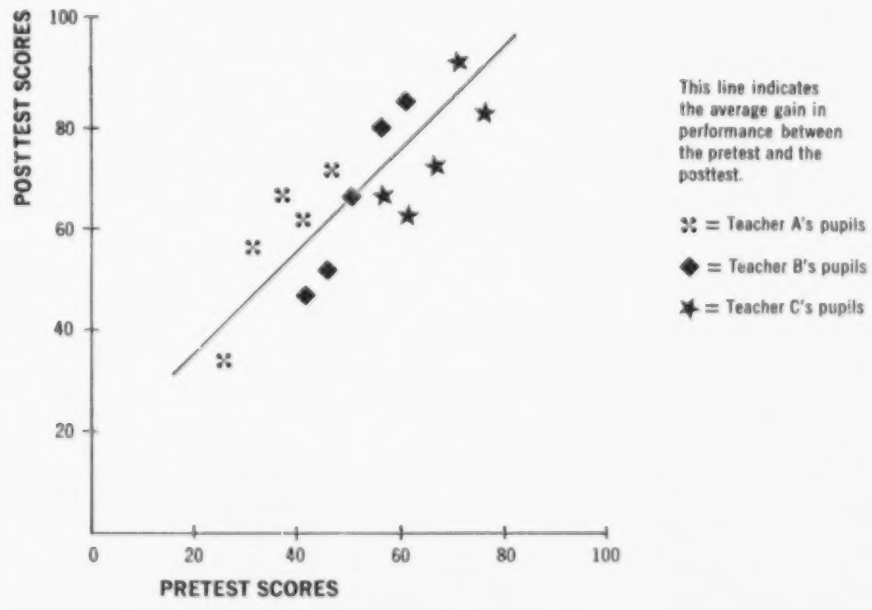


FIGURE 1

The major advantages of this procedure are that it takes into account the student's skills and knowledge before instruction begins, it is flexible enough (via a technique called multiple regression) to take into account several input factors (such as minority group membership and different instructional programs), and it examines more than just the class's average performance. Its major disadvantage, however, is that it requires that students be measured twice and, thus, might not be applicable for districts that have very high student mobility problems. It is also limited to comparing teachers only on a grade-by-grade basis in elementary school and on a subject-by-subject basis at higher levels.

It is apparent, therefore, that the setting of standards against which to evaluate teacher effectiveness can be a difficult job. To obtain adequate controls for potentially important input factors one must use a large sample of teachers and then wrestle with the question of what constitutes satisfactory performance. The problem is not one that will simply go away by itself. If educators fail to establish satisfactory standards, then alternate procedures will be employed. A school board member once suggested to the authors that the way to apportion teacher salaries is directly on the basis of student performance. This would mean that the highest paid teacher, regardless of experience or education (or students worked with) should be the one whose students are performing the best, and so on down the line until the salaries get so low that the "incompetents" seek employment elsewhere. It is apparent that most teachers would prefer some standard to aim for rather than be forced to comply with arbitrary schemes devised by others.

Further Comments

Before leaving the topic of the evaluation of teacher performance, it is necessary to discuss briefly the question of test security and controls. Since the emphasis of an outcome accountability evaluation is on judging the quality of teacher performance, there can be no substitute for extremely high test security. The confounding of state-wide test scores by such things as unauthorized word lists, so-called "practice tests," and similar devices is proof enough of the importance of security. We feel that it is unfortunate that many state-wide tests have been used to evaluate schools rather than for their more appropriate use of counseling individual students. As noted above, state-wide tests rarely are sensitive to the particular curriculum employed at a given school and thus they are not fair teacher-assessment tools. Given the situation of unfair measurement tactics, it is not surprising that many people have tried to subvert it. The solution to this problem is not to provide better test security for state-wide testing programs, but to provide security for the assessment procedures that should be employed in evaluating teachers and school programs. Further, this does not necessarily mean yearly checkups on all teachers with all their pupils but, rather, a systematic and relatively low-cost approach for gathering reliable and valid information periodically. For such a system to be effective it must, however, like Caesar's wife, be beyond reproach.

As we noted in the first section of this paper, the accountability wave is upon us and it is likely to remain for a long time relative to growing legislative support of it. It is also apparent that the procedures needed to implement accountability systems will require a number of controls, like test security and adjustments for input factors, if these systems are to work effectively in meeting both the spirit and the letter of the legislation. History has taught us, however, that whenever one group (such as school administrators) attemps to place controls over another group (such as teachers), then we can expect a counteraction in order to avoid these controls. One state college in California, for example, is already offering a symposium on how teachers can deal with teacher-evaluation legislation. It is evident, therefore, that if an accountability system is to be effective, it should be a joint effort of teachers and administrators. It appears to us that this will only come about through an emphasis on the improvement of the educational process rather than on global judgments of its overall effectiveness. For example, in the case depicted by Figure 1 and Table 1, one might use the results to find out what techniques Teacher A was using in order to gain her relatively higher performance. Such techniques might then be used more widely throughout the district via a variety of programs in which teachers review and critique each other (Niedermeyer and Klein, 1971). Thus, it is our contention that all the controls needed for an accountability system to work could only be implemented successfully if that system also had some payoff for the teachers as well as the administrators, since both groups must support the system if it is to function effectively.

Summary

We have outlined some of the characteristics we consider necessary to an outcome accountability system for teachers. Such systems should be based on the assessment of student performance with measures that are appropriate for this purpose; if nationally-normed standardized tests are insensitive to the goals of specific programs they should not be used. Similarly, techniques of principal ratings or observations are usually invalid and unreliable. We also mentioned the advantages and limitations of a strategy that might be used for this purpose: namely, objectives-based evaluation systems. We also suggested a potentially intriguing but largely untested means by which teacher performance tests might be used along with objectives-based evaluation systems in determining outcome accountability. Quantitative procedures for establishing standards that would "account for" differences in students that teachers will be instructing were discussed along with the steps needed to implement such a system. And finally, we considered the issue of test security in relation to the broader question of controls and the purposes for which an accountability system might be employed.

References

Barro, S. M. An approach to developing accountability measures for the public schools. *Phi Delta Kappan*, 1970, *52*, 196-205.

CSE elementary school test evaluations. Hoepfner, R., Strickland, G., Stangel, G., Jensen, P., & Patalino, M. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education, 1970.

CSE-ECRC preschool kindergarten test evaluations. Hoepfner, R., Stern, C., & Nummedal, S. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education, 1971.

Klein, S. P. Evaluating tests in terms of the information they provide. *Evaluation Comment*, 1970, 2(2), 1-6.

Klein, S. P. The uses and limitations of standardized tests in meeting the demands for accountability. *Evaluation Comment*, (1971, 2(4), 1-7.

Klein, S. P. An evaluation of New Mexico's educational priorities. Paper presented at the Western Psychological Association Convention, Chicago, April 1972.

Lindman, E. Net-shift analysis for comparing distributions of test scores. Working Paper No. 5, 1968. Center for the Study of Evaluation, University of California, Los Angeles.

McNeil, J. Experienced teachers versus novices in attaining results with young learners. Unpublished manuscript. University of California, Los Angeles, 1972.

Niedermeyer, F., & Klein, S. P. An empirical evaluation of a district's teacher accountability system. An Educational Evaluation Associates Report, Los Angeles, California; 1971.

Office of Economic Opportunity. An experiment in performance contracting: Summary of preliminary results. A press release from the Office of Economic Opportunity, Washington, D. C.; February 1, 1972.

Popham, W. J. *Designing teacher evaluation systems.* Los Angeles: Instructional Objectives Exchange, 1971. (a)

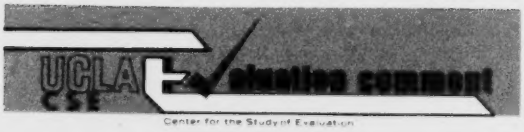
Popham, W. J. Performance tests of teaching proficiency: Rationale, development, and validation. *American Educational Research Journal*, 1971, *8*, 105-117. (b)

EVALUATION COMMENT IS PUBLISHED BY THE CENTER FOR THE STUDY OF EVALUATION, UNIVERSITY OF CALIFORNIA, LOS ANGELES. Each issue of the *Comment* will provide information and ideas about the evaluation of educational programs by presenting articles which discuss evaluation theory, procedures, methodologies, or practices. Periodically, Center products which will facilitate the actual conduct of evaluation studies will be discussed.

A copy of *Evaluation Comment* is distributed free of charge to each scholar, researcher, or practitioner on our mailing list. One to five copies may be obtained free of charge; however, where greater quantities are needed readers are encouraged to reproduce the Comment themselves. To be placed on our mailing list or to order, subject to availability, additional copies of *Evaluation Comment*, please write to:

James Burry, Editor
Evaluation Comment
Center for the Study of Evaluation
145 Moore Hall
University of California, Los Angeles
Los Angeles, California 90024

University of California



Center for the Study of Evaluation
145 Moore Hall
405 Hilgard Avenue
Los Angeles, California 90024

NON-PROFIT ORG.
U.S. POSTAGE
PAID
LOS ANGELES, CALIF.
PERMIT NO. 12378

ERIC CTR FOR TESTS
MEASUREMENT AND EVALUATION
EDUC TESTING SERVICES
PRINCETON, NJ 08540